

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:00:18 ; Search time 22 Seconds

(without alignments)  
450.584 Million cell updates/sec

Title: US-09-887-784-4

Perfect score: 1274

Sequence: 1 MWSKGELFTGVVPIILVELD.....VLLGFTVTAAGITLGMDELK 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1247	97.9	238	1	GFP_AEOVI
2	91.5	7.2	861	1	SYL_HAEIN
3	89.5	7.0	879	1	SYL_XYLEFA
4	87.5	6.9	2222	1	DPOE_YEAST
5	87	6.8	874	1	SLAP_BACLI
6	86.5	6.8	533	1	CP51_CANGA
7	86.5	6.8	795	1	D152_HAEIN
8	86.5	6.8	797	1	D151_HAEIN
9	86	6.8	357	1	TRMA_CAMJE
10	85.5	6.7	788	1	DPOL_HPBHE
11	85.5	6.7	793	1	D153_HAEIN
12	85.5	6.7	886	1	ITH3_MESAU
13	85.5	6.7	941	1	GUN_EAC56
14	85	6.7	439	1	SY62_DISOM
15	84.5	6.6	613	1	PEPF_MYCPU
16	84	6.6	353	1	HIS7_BUCAI
17	84	6.6	504	1	YC03_KLEPN
18	83.5	6.6	538	1	GRBE_RAT
19	83.5	6.6	658	1	ADAS_HUMAN
20	83.5	6.6	1164	1	BAG_STRAG
21	83	6.5	461	1	PSBC_CYPAP
22	83	6.5	774	1	AMY2_SCHPO
23	82.5	6.5	533	1	NIFD_CLOPA
24	82	6.4	682	1	PRC_ECOLI
25	82	6.4	752	1	NECL_RAT
26	81	6.4	336	1	YD48_METJA
27	81	6.4	1224	1	COPA_HUMAN
28	80.5	6.3	393	1	MT04_HELPY
29	80.5	6.3	658	1	ADAS_CAVPO
30	80.5	6.3	860	1	SYL_ECOLI
31	80	6.3	461	1	PSBC_CHLEU
32	80	6.3	737	1	OPT1_DROME
33	79.5	6.2	312	1	TRXB_CHLMU

34	79.5	6.2	468	1	GLNA_AZOCA
35	79.5	6.2	546	1	AMVB_BACCE
36	79.5	6.2	700	1	CAN2_HUMAN
37	79.5	6.2	887	1	ITH3_RAT
38	79	6.2	431	1	TIG_CLOAB
39	79	6.2	836	1	DPOL_HPBDO
40	79	6.2	953	1	YNM7_YEAST
41	78.5	6.2	355	1	PLK_CHICK
42	78.5	6.2	1259	1	YTFN_ECOLI
43	78	6.1	728	1	CATB_ASPFU
44	78	6.1	1148	1	MED_ECOLI
45	78	6.1	1224	1	COPA_BOVIN

P94126	azorhizobiu
P36924	bacillus ce
P17655	homo sapien
O63416	rattus norv
Q97ft6	clostridium
P03162	duck hepati
P53917	saccharomyc
P07354	gallus gall
P39321	escherichia
O92405	aspergillus
P30958	escherichia
Q27954	bos taurus

## ALIGNMENTS

### RESULT 1

ID	GFP_AEOVI	STANDARD;	PRT;	238 AA.
AC	P42212; Q17104;			
DT	Q1-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G., Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent protein.";			
RT	Gene 111:229-233(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
RN	[3]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea green fluorescent protein.";			
RL	Biochemistry 32:1212-1218(1993).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormos M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y., Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent protein.";			
RL	Science 273:1392-1395(1996).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.			
RX	MEDLINE=98455509; PubMed=9782051;			
RA	Wachter R.M., Elsiger M.A., Kallio K., Hanson G.T., Remington S.J.;			
RT	"Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein.";			
RL	Structure 6:1267-1277(1998).			

[7] X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 MEDLINE-99238303; PubMed-10220315;  
 Ellisager M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;  
 "Structural and spectral response of green fluorescent protein  
 variants to changes in pH".  
 Biochemistry 38:5296-5301(1999).  
 CC -1- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE  
 BLUE CHEMILUMINESCENCE OF THE PROTEIN Aequorin INTO GREEN  
 FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON  
 RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN Aequorin.  
 CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER  
 CC ABSORBANCE PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM  
 CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: PHOTOCYTES.  
 CC -1- PTM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED  
 CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON  
 CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.  
 CC -1- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
 CC chimeric proteins of GFP linked to other proteins where it  
 CC functions as a fluorescent protein tag. GFP tolerates N-and C-  
 CC terminal fusion to a broad variety of proteins. It has been  
 CC expressed in bacteria, yeast, slime mold, plants, drosophila,  
 CC zebrafish, and in mammalian cells. As a noninvasive fluorescent  
 CC marker in living cells, it allows for a wide range of applications  
 CC where it may function as a cell lineage tracer, reporter of gene  
 CC expression, or as a measure of protein-protein interactions.  
 CC -1- DATABASE: NAME-Protein Spotlight;  
 CC NOTE-Issue 11 of June 2001;  
 CC WWW="http://www.expasy.org/spotlight/articles/spt011.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; MG2654; AAA27722.1; -;  
 CC EMBL; MG2653; AAA27721.1; -;  
 CC EMBL; L29345; AAA58246.1; -;  
 CC PDB; 1GFL; 11-JAN-97.  
 CC PDB; 1EMA; 08-NOV-96  
 CC PDB; 1EMB; 16-JUN-97.  
 CC PDB; 1EMC; 20-AUG-97.  
 CC PDB; 2EMD; 20-AUG-97.  
 CC PDB; 1EME; 20-AUG-97.  
 CC PDB; 1EMF; 20-AUG-97.  
 CC PDB; 1EMG; 12-MAY-99.  
 CC PDB; 1EMK; 20-AUG-97.  
 CC PDB; 1EML; 20-AUG-97.  
 CC PDB; 1EMM; 20-AUG-97.  
 CC PDB; 2EMN; 20-AUG-97.  
 CC PDB; 2EMO; 20-AUG-97.  
 CC PDB; 1BFP; 07-JUL-97.  
 CC PDB; 1YFP; 28-OCT-98.  
 CC InterPro: IPR000786; Green\_fl\_protein.  
 CC Pfam: PF01353; GFP. 1  
 CC PRINTS; PR01229; GFPLORESCENT.  
 CC ProDom; PD013756; Green\_fl\_protein; 1.  
 CC Luminescence; 3d-structure.  
 KW SITE 65 67 MODIFIED TO FORM THE CHROMOPHORE.  
 FT VARIANT 100 100 F -> Y.  
 FT VARIANT 108 108 T -> S.  
 FT VARIANT 141 141 L -> M.  
 FT VARIANT 219 219 V -> I.  
 FT VARIANT 225 225 H -> Q (IN REF. 2).  
 FT CONFLICT 157 157 Q -> P (IN REF. 2).  
 FT CONFLICT 172 172 E -> K (IN REF. 2).  
 FT SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6E05 CRC64;

Query Match 97.9%; Score 1247; DB 1; Length 238;  
 Best Local Similarity 98.3%; Pred. No. 1.le-99;  
 Matches 234; Conservative 1; Mismatches 3; Indels 0; Caps 0;  
 OY 2 VSKGEELFTGVVPIVLVDGVDNGHGFVSQSGEGDGYGKLTAKFICTTGKLPVWPPTL 61  
 DB 1 MSKGEELFTGVVPIVLVDGVDNGHGFVSQSGEGDGYGKLTAKFICTTGKLPVWPPTL 60  
 OY 62 VTTLSYGQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDNGYKTRAEVKEGDTLV 121  
 DB 61 VTTLSYGQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDNGYKTRAEVKEGDTLV 120  
 OY 122 NRIELKGIDFEDGNILGHKLEYNNSHVYIMADKQNGIKVFNIRINIEDGSVOLAD 181  
 DB 121 NRIELKGIDFEDGNILGHKLEYNNSHVYIMADKQNGIKVFNIRINIEDGSVOLAD 180  
 OY 182 HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLGVLTAAAGITLGMDELYK 239  
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQKDPNEKRDHMYLLGVLTAAAGITLGMDELYK 238  
 RESULT 2  
 SYL\_HABIN  
 ID SYL\_HABIN STANDARD; PRT: 861 AA.  
 AC P43827;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).  
 GN LEUS OR H10921  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA Kerkavage A.R., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.M.,  
 RA McKenney K., Sutton G., Tomb J.-F., Glodek A., Kelley J.M.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +  
 CC diphosphate + L-leucyl-tRNA(Leu).  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U10921;  
 CC TIGR; H32774; AAC22581.1; -;  
 CC InterPro: IPR002302; Leu-trna-syntla.  
 CC InterPro; IPR002300; trna-synt\_1a.  
 CC InterPro; IPR001412; trna-synt\_1.  
 CC Pfam; PF00133; trna-synt\_1; 1.  
 CC PRINTS; PR00985; TRNASYNTHLEU.  
 CC TIGRFAMs; TIGR00396; leus\_bact; 1.  
 CC PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.

KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 FT SITE 42 52 "HIGH" REGION.  
 FT SITE 619 623 "KMSKS" REGION.  
 FT BINDING 622 ATP (BY SIMILARITY).  
 SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;

Query Match 7.2%; Score 91.5; DB 1; Length 861;  
 Best Local Similarity 24.1%; Pred. No. 3.6;  
 Matches 46; Conservative 26; Mismatches 60; Indels 59; Gaps 11;

QY 50 TTGKLPVPWPTLVTLSYGVCFSRYPDHMKQHDFFKSAMPGYVOERTIFFKD-----103  
 DB 314 TGDKLPV-VWAFVLMHYGTGAVMAVPAH-DQDFE-----EFAQKYSLPKQVIAPLA 364  
 QY 104 DQNYKTRAEVKEGDTLVNRIELKGDREKGNILGHKLEYNNSHNVIMADK-OKNGI 162  
 DB 365 DEEIDLTKQAFVEHGLKLVNSDEFGKNF--DGAENG-----IADKLEKLGV 408  
 QY 163 ---KYNFKIRH-----NIEDSGVQLADHYQOONTPIGDDGPVLLPDNHYL- 202  
 DB 409 GRQVNYRLRDWGSVRQRYWGAPIMLTLENGDVVPA-----PMEDLPILPEDVVD 461  
 QY 203 STQSALSADPN 213  
 DB 462 GVKSPINADPN 472

RESULT 3  
 SYL\_XYLFA STANDARD; PRT; 879 AA.  
 ID Q9PBG8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine-tRNA ligase) (Leurs).  
 GN LEUS OR XF2176.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=9a5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Honeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madalena A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.F.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";  
 RL Nature 406:151-159(2000).

CC -|- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) - AMP +  
 CC diphosphate + L-leucyl-tRNA(Leu).  
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -|- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE004031; AAF84975.1; ALT\_INIT.  
 CC InterPro: IPR002302; Leu-trNASyntla.  
 CC InterPro: IPR002300; tRNA-synt\_1a.  
 CC InterPro: IPR001412; tRNA-synt\_1.  
 CC Pfam: PF00133; tRNA-synt\_1.1.  
 CC PRINTS: PRO0985; TRNASYNTHLEU.  
 CC TIGRFAMs: TIGR00396; leuS\_Dact.1.  
 CC PROSITE: PS00178; AA\_TRNA\_LIGASE\_I.1.  
 CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 FT SITE 45 55 "HIGH" REGION.  
 FT SITE 637 641 "KMSKS" REGION.  
 FT BINDING 640 640 ATP (BY SIMILARITY).  
 SQ SEQUENCE 879 AA; 99796 MW; 9FDCB992092919E CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 879;  
 Best Local Similarity 22.2%; Pred. No. 5.4;  
 Matches 44; Conservative 29; Mismatches 68; Indels 57; Gaps 10;

QY 50 TTGKLPVPWPTLVTLSYGVCFSRYPDHMKQHDFFKSAMPGYVOERTIFFKDDGNY--107  
 DB 321 TNEQLPV-VWAFVLMHYGTGAVMAVPGHDQDEF--ANKYGLPIROVIALKEPRNDE 377  
 QY 108 -----KTRAEVKEGDTLVNRIELKGDREKGNILGHKLEYNNSHNVYI 153  
 DB 378 STWEPDVWRDWDYADKTR---EFE---LINSAEFDGLDYQDAFEVLAERFE-----421  
 QY 154 MADKQKNG-IRVNFKIRHNIEDSGVQLADHYQOONTPI-----GDGPVLLPDN 199  
 DB 422 ---RQGRGRRVNYLR---DWGYSRQRWYGCPIPIVYCPGAVPVEDQLPVILPEN 474  
 QY 200 -HYLSTQSALSADPN 216  
 DB 475 VAFSGTGSPIKTDPEWRK 492

RESULT 4  
 DPOE\_YEAST STANDARD; PRT; 2222 AA.  
 ID DPOE\_YEAST  
 AC P21951;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA  
 DE polymerase II subunit A).  
 GN POL2 OR DUN2 OR YNL262W OR N0825.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.  
 RP MEDLINE=90381771; PubMed=2169349;  
 RA Morrison A., Araki H., Clark A.B., Hamatake R.K., Sugino A.;  
 RT "A third essential DNA polymerase in S. cerevisiae";  
 RL Cell 62:1143-1151(1990).  
 RN [2]  
 RP SEQUENCE OF 1-2221 FROM N.A.  
 RC STRAIN=S288c / FY1679;  
 RX MEDLINE=96310631; PubMed=8740425;



15-JUN-2002 (Rel. 41, Last annotation update)  
 Cytochrome P450 51 (EC 1.14.14.-) (CYP51) (P450-L1A1) (Sterol 14-  
 alpha-demethylase) (lanosterol 14-alpha demethylase) (P450-14DM).  
 ERG11 OR CYP51.  
 OS Candida glabrata (Yeast) (Torulopsis glabrata).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5478;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2001-15;  
 RX MEDLINE=96161286; PubMed=8593007;  
 RA Geber A., Hitchcock C.A., Swartz J.E., Pullen F.S., Marsden K.E.,  
 RA Kwon-Chung K.J., Bennett J.E.;  
 RT "Deletion of the Candida glabrata ERG3 and ERG11 genes: effect on cell  
 RT viability, cell growth, sterol composition, and antifungal  
 RT susceptibility.";  
 RL Antimicrob. Agents Chemother. 39:2708-2717(1995).  
 RN [2]  
 RP SEQUENCE OF 60-473 FROM N.A.  
 RC STRAIN=ATCC 2001;  
 RX MEDLINE=95081364; PubMed=7989540;  
 RA Burgener-Kairuz P., Zuber J.P., Jaunin P., Buchman T.G., Bille J.,  
 RA Rossier M.;  
 RT "Rapid detection and identification of Candida albicans and  
 RT Torulopsis (Candida) glabrata in clinical specimens by  
 RT species-specific nested PCR amplification of a cytochrome P-450  
 RT lanosterol-alpha-demethylase (L1A1) gene fragment.";  
 RL J. Clin. Microbiol. 32:1902-1907(1994).  
 CC -1- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL WHICH IS  
 CC CRITICAL FOR ERGOSTEROL BIOSYNTHESIS. IT TRANSFORMS LANOSTEROL  
 CC INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL  
 CC (BY SIMILARITY).  
 CC -1- PATHWAY: Ergosterol biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L40389; AAB02329.1; -;  
 DR EMBL; S75389; AAB32679.1; -;  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Electron transport; Oxidoreductase; Monooxygenase; Membrane; Heme;  
 KW Sterol biosynthesis.  
 FT BINDING 472 472 HEME (BY SIMILARITY).  
 FT CONFLICT 64 64 I -> M (IN REF. 2).  
 FT CONFLICT 473 473 I -> T (IN REF. 2).  
 SQ SEQUENCE 533 AA; 61305 MW; A0506C17507E6EF7 CRC64;  
 Query Match 6.8%; Score 86.5; DB 1; Length 533;  
 Best Local Similarity 21.4%; Pred. No. 5.3;  
 Matches 44; Conservative 32; Mismatches 81; Indels 49; Gaps 8;  
 QY 25 GHKFSVS---GEGEGDAYGKLTAFICHTGKLPVPWPPLVTLSTGVQCFSRYPDH--M 79  
 DB 109 GHEEFNAKLADVSAEAYSHL-----TTTFVGKGYVDCPNHRLM 149  
 QY 80 KOHDFEKSAM-PEGVV-----QRTIFFKDDGNVKTAEVFEFGDTLVNRLKLGIDF 131  
 DB 150 EOKFVKGALYKEAFVRVPLLAEEYKYFRNSKKNENNSGIVDMVMSQPM--TIF 207  
 QY 132 KEDGNILGHKLEYNTNSHNVIWADKQKNGIKVNFKIRNINIEDSGVQLADHYQONTIGD 191  
 DB 208 TASRLSLCKEMRDKLDTDFAYLYSLDKGFTPINF-VFPNLPLEHYRKRDKHAQAIS--- 263

QY 192 GPVLLPDNHYLSTQSALSCKDPNEKRD 217  
 DB 264 -----GTYSLIKERRKND 278  
 RESULT 7  
 ID D152\_HAEIN STANDARD; PRT; 795 AA.  
 AC P44935;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)  
 DE (D-15-Ag) (Outer membrane protein D15).  
 GN HI0917.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glöck A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Ghosh C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U32773; AAC22575.1; -;  
 DR TIGR; HI0917; -;  
 DR InterPro; IPR000184; Bac\_surfAg\_D15.  
 DR Pfam; PF01103; Bac\_surfaceAg; 1.  
 KW Antigen; Outer membrane; Signal; Complete proteome.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.  
 SQ SEQUENCE 795 AA; 87478 MW; B85691FC22E6ED44 CRC64;  
 Query Match 6.8%; Score 86.5; DB 1; Length 795;  
 Best Local Similarity 21.9%; Pred. No. 8.7;  
 Matches 48; Conservative 30; Mismatches 78; Indels 63; Gaps 11;  
 QY 65 LSVGVCFSRYPDHMKOHR-----FKSAMPEGVYQE-----RTI 99  
 DB 427 IGYTGESGYSQASVKODNFGTGAAGAAVSIATKNDYGTSVNLGYTEPYFTKDVSLGGNV 486  
 QY 100 FFKDDGNYKTAEVKEFGDTLVNRIELKIDFKEDGNI---LGH-----KLEYNTNS 148  
 DB 487 FFENYDNSKSDTSNRYKTYGNSVTL-GPPVNNNSYVGLGHTYKISNFALEYN--- 542  
 QY 149 HNYIMADKOK-NGIKVNFKIRNINIEDSGVQLADHYQO-----NTPIGDPVLL 196  
 DB 543 RNLYIQSMFKGNGIKTN-----DFDSFGWNYNSLNRGYFTPKGVKASLG-GRVTI 593  
 QY 197 P--DNNHYLSTQSALSCKDPNEKRDHMLLGFVTAAGITLG 233

Db 594 PGSDNKYKLSADVQGFYPLDRDHLWVWSAKASAGYANG 632

RESULT 9

	TRMA_CAMJE	STANDARD;	PRT;    357 AA.
ID	Q9PP92;		
AC	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DE	tRNA (Uracil <sup>-5</sup> )-methyltransferase (EC 2.1.1.35) (tRNA(M <sup>5</sup> -U54))-methylintranseferase (RUMT).		
GN	TRNA OR CJO831C.		
OS	Campylobacter jejuni.		
OC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;		
CC	Campylobacter.		
OX	NCBI_TaxID=197;		
[1]	SEQUENCE FROM N.A.		
RN	STRAIN=NCTC 11168;		
RC	MEDLINE=20150912; PubMed=10688204;		
RA	Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C., Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S., Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W., Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M., Whitehead S., Barrrell B.G.;		
RR	"The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences.";		
RL	Nature 403:665-668(2000);		
CC	-I- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position 54 (M <sup>5</sup> -U54) in all tRNA (By similarity).		
CC	-I- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-homocysteine + tRNA containing thymine.		
CC	-I- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. TRMA SUBFAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).		
DR	EMBL; ALJ39076; CABY3096.l; .		
DR	InterPro; IPR000051; SAM_bind.		
DR	InterPro; IPR001566; TrmA.		
DR	PROSITE; PS01230; TRMA_1; 1.		
DR	PROSITE; PS01231; TRMA_2; FALSE_NEG.		
KW	Transferase; Methyltransferase; tRNA processing; Complete proteome..		
.FT	DOMAIN      207     213       S-ADENOSYLMETHIONINE BINDING (BY SIMILARITY).		
FT	ACT_SITE    315     315           BY SIMILARITY.		
SQ	SEQUENCE    357 AA;    42276 MW;    CEE5328347CEE497 CRC64;		
Query Match	6.8%; Score 86; DB 1; Length 357;		
Best Local Similarity	24.8%; Pred. No. 3.6;		
Matches	Conservative 18; Mismatches 39; Indels 34; Gaps 5;		
QY	80 QKHDFKFSAMPEGVVOERTIFFKDGDGNYKTAEVKF--EGDTLV-----NRLEKG 128 ::    :    :    :                  :		
Df	14 EKHSFIKKYKFEEFYTKDLFAFKDKHRYTRAELSFVHEHTLFYAMPDPKSCKKYIIIEY 73 :    :    :    :                  :		
QY	129 IDPFED-----CNLGHKLEYNNVNSHNVTIMADKKNGIKVNFXIRNHTE 173    :  :  :                      :		
Df	74 LDFADEKICAFPRLLEYLRQNKLKEKL-----FGVEFLTTKQE--LSITLLYHKNIIE 125 :    :    :                      :		
QY	174 D 174		
Df	126 D 126		
RESULT 10			

```

DPOL_HPBHE          STANDARD;          PRT;          788 AA.
ID DPOL_HPBHE
AC P13846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P protein [includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-
DE directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN P.
OS Heron hepatitis b virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88333160; PubMed=3418788;
RA Sprengel R., Kaleta E.F., Will H.;
RT *Isolation and characterization of a hepatitis B virus endemic in
RT herons.;
RL J. Virol. 62:3832-3839(1988).
RC -J- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + [DNA](N).
CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22056; AAA45738.1; -.
DR PIR; A30082; JDVLHH.
DR InterPro; IPR001462; DNAPol_viral_C.
DR InterPro; IPR000201; DNAPol_viral_N.
DR InterPro; IPR000477; Rvtse.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00242; DNAPol_viral_N; 1.
DR Pfam; PF00336; DNAPol_viral_C; 1.
DR ProDom; PD000814; DNAPol_viral_C; 1.
KW Transfrase; RNA-directed DNA polymerase; DNA-directed DNA polymerase;
KW Hydrolyase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SQ SEQUENCE 788 AA; 90070 MW; FB44F3875EADF44 CRC64;

Query Match          6.7%; Score 85.5; DB 1; Length 788;
Best Local Similarity 19.7%; Pred. No. 10;
Matches 40; Conservative 32; Mismatches 66; Indels 65; Gaps 10;

QY 58 WPTLVTTLSYGVCFSRYPDHMKQH-----DFFKSAMPEGYVOERT-----IFFKDDGNYK 108
DB 139 WPKSISYLPVHSGVKPKYPEFQONHESLVNDYLNKLFEGALYKRVSKHLVTFK--GPYF 196
QY 109 T-----RAEVKFEGLTLVNRLEKIDGIDFKEDGNTI---LGH-----KLEYNYS 148
DB 197 TEQKHLVQQHAYSSKINDRQESRRRIITATSSKNDSSRI-----FGAHN----- 245
QY 155 ADKQKGIKVNFKIRHNIEDGSVOLADHYQ-----QNTPIGDGPVLL--PDNHYL 202
DB 246 -----NGRKISY---HSTRDGSRLSGRTSDPTSRGALAGDSTPIGPGSTAAPHSHV 297
QY 203 STQ-----SALSQDPNEKR 216
DB 298 DRRRRQKGGVGLQAISREPSETR 320

RESULT 11
D153_HAEIN          STANDARD;          PRT;          793 AA.
ID D153_HAEIN
AC O32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PAK 12085;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -|- SUBCELLULAR LOCATION: Outer membrane.
CC -|- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U60834; AAB61977.1; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A14 CRC64;

Query Match          6.7%; Score 85.5; DB 1; Length 793;
Best Local Similarity 22.4%; Pred. No. 11;
Matches 49; Conservative 28; Mismatches 79; Indels 63; Gaps 11;

QY 65 LSYGVOCFSRYPDHMKOHDF-----FKSAMPEGYVOE-----RTI 99
DB 427 IGYTSGISYQTSIKQDNFLGTCGAANSIAGTKNDYGVSNLGYTEPFTKDGVSLLGNI 486
QY 100 FFKDDGNYKTRAEVYKFEGLTLVNRLEKIDGIDFKEDGNTI---LGH-----KLEYNYS 148
DB 487 FPEYDNKSDTSSNYKRTTYGNSVTL-GFPVNENNSYVGLGHTYKISNPALEYN--- 542
QY 149 HNYIADKQK-NGIKVNFKIRHNIEDGSVOLADHYQ-----NTPIGDGPVLL 196
DB 543 RNLYIQSMKPKGNGIKTN-----DFDFSFGWNNYSLNRGYPFTKGVKASLG-GRVTI 593
QY 197 P--DNHYLSQTSALSQDPNEKRDRHVLGLFVTAAGITLG 233
DB 594 PGSDNKKYKLSADVQGFYPLDRDRHVVVSASAKSAGYANG 632

RESULT 12
ITH3_MESAU          STANDARD;          PRT;          886 AA.
ID ITH3_MESAU
AC P97280;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inter-alpha-trypsin inhibitor heavy chain H3 precursor (ITI heavy
DE chain H3) (HC3).
GN ITH3.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.

```

[illegible]





```
Matches 44; Conservative 40; Mismatches 75; Indels 29; Gaps 9;
QY 65 LSYGVQCFKSYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLVNRI 124
Db 1 MSKEIKYKYSVPKRYRFDLDYLLEGKTIDQLF---DQFLEKSKKLKIK-DSKYONI 56
QY 125 E--LKGIDFKEDGNILGHKLEYNYSNHYI-MADKQKNGIKVNFKIRH-----NIEDGS 176
Db 57 ESYLESLEEDFNLLNKKI-INYISNNISVNVYDHFREISQKFEFMYYSFFNQIGDEN 115
QY 177 VOLADHYQONTPIGDGPVLLP-----DNHYLSTQSALSKDPNEKRDHMLLVLLGFVTAAGIT 231
Db 116 QRILEHEEKITKLLDPRLASYYKALDFVFKSKHRLSK---EVEDYLI-----KVS 164
QY 232 LGMDELYK 239
Db 165 RGNIELYK 172
```

Search completed: June 3, 2003, 15:07:06  
Job time : 24 secs